THE RELATIONSHIP BETWEEN PARAMETERS FROM TWO POLYTOMOUS ITEM RESPONSE THEORY MODELS

Various polytomous IRT models parameterize the probability of response categories differently from each other. For example, the graded response model (GRM) (Samejima, 1969) is based on the cumulative log-odds principle, whereas the generalized partial credit model (GPCM) (Muraki, 1992) is based on the adjacent log-odds principle. It is widely known that these two polytomous IRT models do not produce directly comparable parameters (e.g., Ostini & Nering, 2005). However, to our knowledge, it has not been clearly pointed out that how discrimination parameters (*a*-parameters, hereafter) from these two polytomous IRT models are affected by the number of response categories. Thus, this study investigates the relationship between the *a*-parameters and the number of response categories for polytomous item response theory models, specifically for the GRM and GPCM. The relationships are first explored algebraically. More specifically, the cumulative category response functions of the investigated models are solved with respect to the *a*-parameter. Then, the algebraically derived relationships are empirically demonstrated with simulated data sets. Finally, practical importance of the findings is discussed.

Method

Algebraic Demonstrations

The formulas for the *a*-parameters based on the cumulative category response functions were derived for the GRM and GPCM. The category response functions $\left(P_{u_i}(\theta)\right)$ of Samejima's graded response model can be expressed as

$$P_{u_i}(\theta) = P^*_{u_i}(\theta) - P^*_{(u_i+1)}(\theta),$$

where $P^*_{u_i}(\theta)$ is the cumulative category response function. On the other hand, since the generalized partial credit model does not belong to the homogeneous case, its cumulative category response function can be expressed as the sum of the category response functions of category u_i and higher (Thissen & Steinberg, 1986). Therefore, the cumulative category response function of the generalized partial credit model is

$$P^*_{u_i}(\theta) = \sum_{u_i=h}^{m} P_{u_i} = \frac{\sum_{h}^{m} \exp\left[\sum_{u_i=1}^{h} Da_i(\theta - b_i + d_{u_i})\right]}{\sum_{c=1}^{m} \exp\left[\sum_{u_i=1}^{c} Da_i(\theta - b_i + d_{u_i})\right]}.$$

From here, these equations are solved with respect to *a*-parameters, such that factors that affect the magnitude of *a*-parameters can be evaluated.

Numerical Demonstrations

After the algebraic derivations, the differences between *a*-parameters from the two models are demonstrated numerically with simulated data. Polytomous item response data sets for 20 items and 5,000 examinees were generated by the GRM, as well as by the GPCM. This way, we can evaluate how a-parameters are affected when data generation models are not the same as the parameter estimation model for both GRM and GPCM. Demonstrations include cases with 4 different numbers of response categories (2, 3, 4, and 5).

Results

First, it has been algebraically demonstrated that the *a*-parameter for the GRM and the GPCM are the same when there are only two response categories. Second, it has been algebraically shown that the derived formula for the *a*-parameter for the GRM remains the same for any number of response categories. Third, however, the derived *a*-parameter formula for the GPCM was quite different for three or more response categories. More specifically, it was

revealed that the item discrimination parameter of the generalized partial credit model depends on (1) the cumulative category response function ($P_{u_i}^{*}(\theta)$), (2) the scaling constant (*D*), and (3) the category parameter (*d*).

Under the assumption that the GPCM and the GRM fit equally well to a given set of item response data, it can be assumed that the cumulative category response functions ($P_{u_i}^{*}(\theta)$) are common between the GPCM and the GRM. Also, *D* is just a constant number (1.7). Thus, the factor that makes the *a*-parameter different between GPCM and GRM is the category parameter (*d*), which in fact depends on the number of response categories. Specifically, it has been algebraically revealed that having more response categories for an item leads to lower *a*-parameter value in the GPCM.

Importance of the Study

The cumulative category response functions of the models that belong to the homogeneous case such as the cumulative category response functions of the GRM are identical in shape (Samejima, 1969; 1995). Therefore, it makes sense that the item discrimination parameter of the graded response model does not depend on the number of response categories. However, the cumulative category response functions of the models that belong to the heterogeneous case are not identical. Therefore, it also makes sense that we obtained the result that indicates the *a*-parameter of the GPCM depends on the number of response categories.

The results of this study are practically important because many large-scale testing programs have an item pool consisted of items with different numbers of response categories, which is also referred to as a mixed item format. In such as case, it is common to calibrate item parameters by utilizing a combination of an IRT model for dichotomous items (such as 2PL and 3PL IRT models) and a polytomous IRT model (such as the GPCM). Calibrated item parameters

are then typically stored in an item bank, and *a*-parameter values are readily available to educational measurement professionals for the purpose of item selections and test construction. However, this study has revealed that we cannot compare estimated *a*-parameters from dichotomous IRT model and GPCM directly, or even between *a*-parameters derived from the GPCM unless items have the same number of response categories. In other words, the *a*-parameters from the GPCM should not be interpreted as a measure of item discrimination power in a test or an item pool, if the number of response categories vary from item to item. Therefore, it will not be a correct item selection strategy to choose an item based on its *a*-parameter value. One can easily overlook a polytomous item with high discrimination power, which may be associated with a low *a*-parameter value simply because of its many response categories. Instead, one probably should evaluate item parameter estimate values in a more global form, such as response category characteristic curves and item information function.

References

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159-76.

Ostini, R., & Nering, M. L. (2005). *Polytomous item response theory models*. Thousand Oaks: Sage.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, Monograph Supplement*, *34*, 100–114.

Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, *60*(4), 549-72.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51*, 567-577.